Cardiology in the Young

cambridge.org/cty

Data quality methods through remote source data verification auditing: results from the Congenital Cardiac Research Collaborative

Joelle A. Pettus[1], Amy L. Pajk[2], Andrew C. Glatz[3], Christopher J. Petit[1], Bryan H. Goldstein[4], Athar M. Qureshi[5], George T. Nicholson[6], Jeffery J. Meadows[7], Jeffrey D. Zampi[8], Mark A. Law[9], Shabana Shahanavaz[10], Michael S. Kelleman[1], Courtney M. McCracken[1] and the Congenital Cardiac Research Collaborative

[1]Emory University School of Medicine, Department of Pediatrics, Children's Healthcare of Atlanta, Atlanta, GA, USA; [2]The Heart Institute, Cincinnati Children's Hospital and Department of Pediatrics, Cincinnati, OH, USA; [3]The Cardiac Center, Department of Pediatrics, Children's Hospital of Philadelphia, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA; [4]Heart Institute, UPMC Children's Hospital of Pittsburgh and Department of Pediatrics, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA; [5]The Lillie Frank Abercrombie Section of Cardiology, Department of Pediatrics, Texas Children's Hospital, Baylor College of Medicine, Houston, TX, USA; [6]Division of Cardiology, Department of Pediatrics, Vanderbilt University School of Medicine, Nashville, TN, USA; [7]Division of Cardiology, Department of Pediatrics, University of California San Francisco School of Medicine, San Francisco, CA, USA; [8]Division of Cardiology, Department of Pediatrics, CS Mott Children's Hospital, University of Michigan School of Medicine, Ann Arbor, MI, USA; [9]Division of Pediatric Cardiology, Department of Pediatrics, Children's of Alabama, University of Alabama Birmingham School of Medicine, Birmingham, AL, USA and [10]Section of Pediatric Cardiology, Department of Pediatrics, Washington University School of Medicine, St Louis, MO, USA

Original Article

Cite this article: Pettus JA, Pajk AL, Glatz AC, Petit CJ, Goldstein BH, Qureshi AM, Nicholson GT, Meadows JJ, Zampi JD, Law MA, Shahanavaz S, Kelleman MS, and McCracken CM (2021) Data quality methods through remote source data verification auditing: results from the Congenital Cardiac Research Collaborative. *Cardiology in the Young* **31**: 1829–1834. doi: 10.1017/S1047951121000974

Received: 3 November 2020
Revised: 19 February 2021
Accepted: 22 February 2021
First published online: 17 March 2021

Keywords:
Audit; database; quality; cardiac; remote

Author for correspondence: Joelle A. Pettus, MPH, MSW, Emory University School of Medicine, Department of Pediatrics, 2015 Uppergate Drive, Atlanta, GA 30322, USA. Tel: 404-727-5198; Fax: 770-488-9015. E-mail: JPettus@emory.edu

Abstract

Background: Multicentre research databases can provide insights into healthcare processes to improve outcomes and make practice recommendations for novel approaches. Effective audits can establish a framework for reporting research efforts, ensuring accurate reporting, and spearheading quality improvement. Although a variety of data auditing models and standards exist, barriers to effective auditing including costs, regulatory requirements, travel, and design complexity must be considered. *Materials and methods:* The Congenital Cardiac Research Collaborative conducted a virtual data training initiative and remote source data verification audit on a retrospective multicentre dataset. CCRC investigators across nine institutions were trained to extract and enter data into a robust dataset on patients with tetralogy of Fallot who required neonatal intervention. Centres provided de-identified source files for a randomised 10% patient sample audit. Key auditing variables, discrepancy types, and severity levels were analysed across two study groups, primary repair and staged repair. *Results:* Of the total 572 study patients, data from 58 patients (31 staged repairs and 27 primary repairs) were source data verified. Amongst the 1790 variables audited, 45 discrepancies were discovered, resulting in an overall accuracy rate of 97.5%. High accuracy rates were consistent across all CCRC institutions ranging from 94.6% to 99.4% and were reported for both minor (1.5%) and major discrepancies type classifications (1.1%). *Conclusion:* Findings indicate that implementing a virtual multicentre training initiative and remote source data verification audit can identify data quality concerns and produce a reliable, high-quality dataset. Remote auditing capacity is especially important during the current COVID-19 pandemic.

Complete, accurate, and verifiable data are essential to the validity and reliability of any research dataset. Absence of a well-defined data dictionary and specific methods to measure discrepancy undermine data accuracy, reliability, and quality. Thus, rigorous, consistent, and objective data auditing and verification processes are necessary core competencies for an effective and reliable dataset that aims to advance research and improve healthcare outcomes.

The Congenital Cardiac Research Collaborative is a multicentre paediatric cardiology research consortium that aims to produce high-quality, generalisable outcomes research to address significant clinical questions in the management of CHD.[1] Through the use of multicentre retrospective data collection, the CCRC has studied several lesions and treatment strategies in CHD including congenital aortic valve stenosis, pulmonary atresia with an intact ventricular septum, and ductal-dependent pulmonary blood flow.[2-4] Following an expansion in site membership in 2018 to nine centres in the United States of America, the CCRC embarked on its fourth and largest major study comparing outcomes in neonates with symptomatic

© The Author(s), 2021. Published by Cambridge University Press.

CAMBRIDGE UNIVERSITY PRESS

tetralogy of Fallot who underwent either staged repair (initial palliation followed by a subsequent complete repair) or primary repair in infancy (the Infant Tetralogy of Fallot cohort study).[5]

While detailed data auditing measures have been successfully employed for all CCRC projects to date, the collaborative initiated a more ambitious and systematic approach to assessing data quality and accuracy for the Infant Tetralogy of Fallot cohort study by utilising source data verification. A source data verification audit is the practice of verifying data in a database, dataset, or registry against primary source documentation. This practice can be completed manually or electronically and conducted either on-site or remotely. A source data verification audit allows investigators an opportunity to verify entered data to the database source, identify repetitive or systematic issues with data collection, and gain confidence in making inferences on study findings.[6]

Data collection for the Infant Tetralogy of Fallot cohort study required extensive retrospective chart review across the nine participating CCRC member institutions. A detailed remote data auditing and quality testing initiative was executed amongst the collaborative institutions. This paper aims to describe the methodology and results of the CCRC verification audit to provide justification for the quality and integrity of both the CCRC's data training and auditing initiatives and the Infant Tetralogy of Fallot cohort study dataset. Further, we explore the unique value of remote source data verification audits during the current coronavirus disease 2019 pandemic.

Materials and methods

Regulatory structure

Cincinnati Children's Hospital Medical Center served as the single institutional review board of record for the Infant Tetralogy of Fallot cohort study with reliance agreements established amongst the other eight participating centres. Children's Healthcare of Atlanta, along with Emory University, served as Data Coordinating Center. Data use agreements between each CCRC institution and the Data Coordinating Center were established. A study leadership team was comprised of the two overall study Principal Investigators, the CCRC Biostatistics Chair, Data Coordinating Center staff biostatistician, the CCRC program manager, lead study research coordinator, and the CCRC Scientific Committee Chair.

Data collection training

Paper case report forms derived from the study protocol were developed for two distinct treatment arms, staged repair and primary repair. Following extensive editing and beta testing, the case report forms were translated into electronic instruments using the Research Electronic Data Capture database hosted at the Children's Healthcare of Atlanta. REDCap is a HIPAA-compliant web-based data capturing tool that allows for robust data capture across institutions.[7] The CCRC utilised REDCap features such as complex branching logic, required fields, drop-down menus with predefined choices, minimum and maximum set values, and date verification rules to minimise potential human error and clean the data from the beginning steps of data entry. Further, routine verification checks to reduce inconsistencies, range, and logic discrepancies and eliminate nonsense values, unnecessary repetitions, and inconsistent or egregious dates were implemented throughout the study.[8]

Prior to finalising the database, 30 team members across 9 sites were identified and assigned to 1 of the 3 roles: data extraction, data entry, or both. Data extractors were responsible for extracting data from the local electronic medical record used in completing the case report forms; data entry personnel were only permitted to transfer data from completed case report forms to the REDCap database. Study staff permitted to both extract and enter data were able to enter data directly into REDCap without completing the paper case report forms.

A manual of operations, standard operating procedure checklists, and a review of study-specific data dictionary elements were crafted and circulated. All team members were required to attend a 1-hour interactive virtual training session. This session included a study overview, an explanation of all study materials, and tutorials on navigating REDCap, the CCRC data training process, medical chart review, and data auditing. To test for data transfer and entry accuracy, data entry personnel were sent three simulated test cases in paper case report form and instructed to enter the data into phantom patient records in REDCap. If two or more discrepancies were found on any one REDCap form, the study staff were required to retest with another simulated dataset. No institution or individual required formalised data entry retraining.

Data extractors underwent an internal medical chart review and extraction training with their institution's Principal Investigator. Data extraction testing consisted of the site Principal Investigator and each data extractor completing a full set of case report forms for the same three records: one staged repair with initial transcatheter-based palliation, one staged repair with initial surgical palliation, and one primary repair. Principal Investigator completed case report forms served as the gold standard and all discrepancies were recorded and shared with the Principal Investigator to provide additional site training as needed. Common discrepancies across sites were reviewed and communicated to the team. While discrepancies were discovered, no site or data extraction individual required additional, formalised retraining as data extraction standards were met across all sites. Finally, data were collected between 12 December, 2018 and 15 April, 2019 for all eligible TOF patients whose initial intervention occurred between 1 January, 2005 and 30 November, 2017.

Planning and procedure of data audit

Prior to any official auditing, the Data Coordinating Center implemented periodic validation checks to assess initial data completeness, biological plausibility, chronological flow, and anomalous data points. Next, three distinct types of queries were conducted for an initial data assessment to remove (1) outliers, (2) ineligible patients, and (3) missing or incomplete data. Egregious ages, dates, and measurement outliers were identified using descriptive statistics, graphical plots, and REDCap visual reports. All queries were reviewed, adjudicated by the site team, and updated in the database. *A priori*, study leadership established a target overall discrepancy rate for key data elements of <5%.

Source data verification audit

Based on the planned primary and secondary outcomes and key covariates of interest, the study leadership team identified 51 key data fields across both groups for the source data verification audit (Supplementary Table 1). Two factors led to differences

https://doi.org/10.1017/S1047951121000974 Published online by Cambridge University Press

in the number of variables audited per patient. First, the distinct differences between the staged repair and primary repair groups led to differences in the number of key variables per medical record. For example, the mean number of key variables for the primary repair group was 28 and 33 for the staged repair group. Second, records that included multiple reinterventions contributed more key data variables across both groups. The maximum number of possible key variables per patient with one reintervention per strategy was: 33 for primary repair, 44 for catheter palliation, and 39 for surgical palliation.

The Data Coordinating Center generated a 10% (5 minimum–10 maximum per site) random sample of patient records selected to audit. Each institution identified a source data verification auditing representative responsible for gathering all relevant medical records for data verification, de-identifying all files, and uploading them to the secure REDCap portal. These individuals received additional training on source documentation uploading and de-identification techniques to remove protected health information. A standard operating procedure document and video tutorial were circulated instructing users to redact all prohibited protected health information using Adobe tools and review with the Principal Investigator and another study staff member before uploading. Any instances of protected health information were reported to the site and central Institutional Review Board. The files were removed, and re-education was implemented as needed. If any site failed to meet the established passing score of 95% accuracy or higher, the site would be required to complete an additional audit with a new set of randomly selected records.

CCRC institutions de-identified and uploaded the source auditing materials within 33 days with a 100% response rate. Two trained data auditors, the project manager and lead research coordinator, extracted data from the uploaded, de-identified files, and entered the key variables into mirrored REDCap records. While both auditors were aware of which site these files were associated with, they were blinded to the treatment plan and to the original records while auditing. That is, the auditors entered all key variables as if completing an entirely new patient and then compared the two sets of entered data, the original records from the site and the auditor-completed records, using REDCap's data comparison tool.

Discrepancy classification

Discrepancies were divided into three subcategories – transcription/clerical, inaccurate reading, and inaccurate assessment per protocol. A transcription/clerical discrepancy included typographical errors and accidental inversions, while common examples of inaccurate reading consisted of dates inaccurately reported by less than 2 days of the actual date listed on the medical record. Inaccurate assessments per protocol included more severe accurate readings that could skew data such as missing a procedure on a "check all" question or omitting a procedural complication (Supplementary Table 2).

Additionally, each discrepancy was scored on the severity of the error, either minor or major. Minor discrepancies were defined as imperfect matches between the data submitted by the site and the auditor's adjudication where the discrepancy would not be expected to significantly alter data analysis or interpretation of results. For example, inaccurate dates within 2 days of the correct date listed and clerical errors where a weight was entered inaccurately by less than one unit were considered minor. Major discrepancies were defined as instances in which discrepancies led to inaccurate assessments per study protocol or altered data in a manner that could potentially lead to significantly skewed or incorrect results. For instance, dates entered that were 3 or more days off from the date listed in the medical record or missing complications or reinterventions were classified as major discrepancies (Supplementary Table 2).

The auditors requested verification on any unresolved concerns during site-specific conference calls. Discrepancies were reported and reviewed with the study team who verified if a true discrepancy occurred or a discrepancy in abstraction was made by either the site data verification auditing representative or the auditor. Systemic, reoccurring discrepancies were noted and reviewed with the study team during weekly calls and additional training and clarification were provided as necessary.

All key data fields were included in the analysis; however, duplicated discrepancies did not contribute to scoring criteria. That is, if a discrepancy was made on a key variable with a child–parent relationship to other variables, the "child" fields were not included in scoring criteria.[9] For instance, if a reintervention visit was missed and listed as "not applicable", the remaining key variables on the reintervention instrument were not counted as additional, duplicated discrepancies.

Statistical methods

Standard descriptive statistics were used to summarise the counts and percentages of discrepancies, both overall and by type. Differences in the proportion of discrepancies between centres and treatment groups were assessed by Pearson chi-square test. Statistical significance was established *a priori* at a two-tailed alpha level of 0.05. All analyses were performed using STATA v10 (Stata Corp., College Station, TX, USA) or higher.

Results

Out of a total of 572 study patients, 58 randomly selected patients (27 primary repairs and 31 staged repairs) were audited for all available key data fields depending on the patient group. In total, 1790 data points were audited (mean 30.9 ± 5.3 variables per patient). Amongst the 1790 data points, a total of 45 discrepancies were discovered, resulting in an overall accuracy rate of 97.5% with a discrepancy rate of 2.5% (Table 1). At the institution level, data accuracy ranged from 94.56% to 99.36% (Table 1). Of the total 1790 variables, 951 or 53% were from staged repair records while 839 or 47% were derived from primary repair records. Of the total 45 discrepancies, 24 were found in the staged repair records, representing an overall discrepancy rate of 2.5% of all audited staged repair data fields. Twenty-one discrepancies were discovered in the primary repair group, representing a group discrepancy rate of 2.5%. There was no statistical difference in overall discrepancy rates between the two study groups (Table 2) ($p = 0.98$). Amongst the 58 randomly selected patients, 27 patients had 0 data errors, 18 patients had 1 data error, 12 patients had 2 errors, and 1 patient had 3 or more errors. The average accuracy rate per patient was 97.4 ± 2.8%.

Of the total 45 discrepancies, 26 (58%) were classified as minor discrepancies, for an overall minor discrepancy rate of 1.5% (Table 3). Common instances of minor discrepancies included dates entered inaccurately within 2 days of the correct date, and instances in which all data were included but listed in the incorrect place in the database. For example, a minor discrepancy involved listing a complication under the "other" free text slot

https://doi.org/10.1017/S1047951121000974 Published online by Cambridge University Press

Table 1. Key findings

Site	Medical records audited	Total variables	Total discrepancies	Per cent accuracy	P-value	Records with ≥ 1 discrepancy	Records with ≥1 major discrepancy
1	5	163	5	96.93%	0.045	3	2
2	8	253	4	98.42%		4	1
3	7	223	6	97.31%		4	3
4	10	311	2	99.36%		2	2
5	5	155	1	99.35%		1	0
6	5	147	8	94.56%		5	3
7	5	155	6	96.13%		5	2
8	6	204	8	96.08%		4	2
9	7	179	5	97.21%		3	2
Total	58	1790	45	97.5%		31	17 (29.3%)

Table 2. Discrepancies between cohorts

Staged repair versus primary repair					
Cohort	Total variables	Total discrepancies (%)	P-value	Major discrepancy rate	P-value
Primary repair	839	21 (2.5%)	0.98	10 (1.2%)	0.98
Staged repair (all)	951	24 (2.5%)		9 (0.9%)	
Total	1790	45		19	

Table 3. Discrepancy classification

Major versus minor discrepancy classification			
Discrepancy classification	Classification count	Percentage to total discrepancies	Discrepancy rate
Major	19	42%	1.1%
Major primary repair – 10 *Major staged repair (all) – 9*			
Minor	26	58%	1.5%
Minor primary repair – 11 *Minor staged repair (all) – 15*			
Total discrepancies	45		
Total variables	1790		

Table 4. Discrepancy type

Discrepancy type breakdown			
Discrepancy type	Discrepancy count	Per cent discrepancy type to total discrepancies	Per cent discrepancy type to total auditing dataset
Transcription/ clerical	25	56%	1.4%
Inaccurate read – days	11	24%	0.6%
Inaccurate assessment per protocol	9	20%	0.5%
Total discrepancies	45		
Total variables	1790		

when a provided answer choice should have been selected. The remaining 19 discrepancies (42%) were classified as major discrepancies, for an overall major discrepancy rate of 1.1%. The 19 major discrepancies were discovered amongst 17 patients across the total 58 individual patient records audited, for a per-patient 29.3% major discrepancy rate (Table 1).

The most common major discrepancy was instances in which dates deviated by 3 days or more from the correct date in the medical chart. The auditors also identified occurrences of missed procedural complications or missed or inaccurate details from complex reintervention encounters (Supplementary Table 2). Transcription/clerical errors consisted of 56% of all discrepancies and 1.4% of all variables in the aggregate auditing dataset. Inaccurate reading of days garnered 24% of total discrepancies and 0.61% of the aggregate auditing dataset. Inaccurate assessment per-protocol discrepancies, arguably the most concerning type constituted 20% of all discrepancies and only 0.5% of all auditing data variables gathered (Table 4).

https://doi.org/10.1017/S1047951121000974 Published online by Cambridge University Press

Discussion

We report the processes and results of the remote source data verification audit for the CCRC Infant Tetralogy of Fallot cohort study, with an overall accuracy rate for the key data fields of 97.5% and a major discrepancy rate of only 1.1%. With institutional discrepancy rates ranging from 94.6% to 99.4%, no significant source-to-database discrepancy rate differences were found amongst sites. While a < 10% discrepancy rate is generally acknowledged as an adequate standard, inconsistent guidelines still exist regarding "acceptable" discrepancy rates.[10]

The CCRC established and successfully met its internal guideline of <5% discrepancy rate for the overall data auditing cohort. Not only do these findings have important implications for the successful generation of a highly reliable multicentre dataset, but these processes may also be particularly relevant during the current COVID-19 pandemic where the ability to travel for on-site audits is limited. Given the success, these strategies could also be used as a basis for a permanent model of remote data auditing that effectively utilises resources.

While the presence and expectation of a thorough auditing mechanism is not a new concept in medical research, there are multiple methods, procedures, and tools utilised in medical research to ensure data quality that may vary by category. Clinical trials, in particular, are held to high standards for accuracy, completion, and execution. Large-scale registries, whether clinically based or quality improvement focused, must implement effective strategies to gather clean, accurate data from the onset. In the cardiac field, high data accuracy rates have been presented within a variety of registries, consortiums, and datasets – both retrospective and prospective. The National Cardiovascular Data Registry Data Quality Program reported data abstraction accuracy rates at 93.1, 91.2, and 89.7% for three large-scale cardiovascular registries.[11] The Pediatric Cardiac Critical Care Consortium reported an overall accuracy of 99.1%[12] while The Society of Thoracic Surgeons Congenital Heart Surgery described an aggregate 97.4% rate of agreement in a recent 10-year data audit review.[13] Significantly, such registry datasets often select variables in advance and are aware of what data will be extracted before any collection methods begin. Working retrospectively, as the CCRC did here, adds an additional layer of complexity and requires an understanding of the variety of available data across multiple sites.

Although clinical trials and large-scale registries in the cardiac field and elsewhere have been conducting remote data auditing and source verification for some time,[14,15] limited literature exists on the assessment of non-automated source data verification audits for retrospective chart review studies and their ability to detect random and systematic discrepancies.[10] Further, smaller registries or single datasets are typically not held to the same rigorous standards for robust auditing and remote data capture as clinical trials. While in-person audits are often considered the preferred method for increased consistency and rigour, it is not always feasible to conduct in-person audits for multisite research. Additionally, remote audits are more cost-effective, time-efficient and allow for increased flexibility and deadline adherence. For these reasons, the CCRC developed and applied manual methods to perform an internal source data verification audit.

Though similar, excellent auditing processes exist and have been reported for many quality improvement or clinical datasets, to our knowledge no detailed reporting of auditing processes for a multicentre retrospective research dataset in CHD has been previously described. Thus, we believe that this manual, remote, yet comprehensive auditing initiative presents unique findings and alternative approaches that similar collectives can adapt to ensure high-quality data. The results of this initial audit confirm that manual, remote source data verification audits can produce effective, efficient non-automated audit procedures, and standards across multisite retrospective research studies while reducing costs, time, and travel. Such methods can be adopted as an alternative to adapt to workforce changes due to the COVID-19 pandemic and potential long-term modifications that may result in research administration, execution, and auditing efforts.

Limitations

Much knowledge was gathered throughout this inaugural source data verification audit process, which allowed the opportunity to identify weaknesses within our data training, collection, and storing process and recommendations to resolve issues for future data quality efforts. One notable limitation experienced in this process was the need for the auditors to request additional documentation or the location of specific source data. If a site is omitted or neglected to provide files, particularly for more complex hospitalisations, the auditors could be left unaware. This constraint was negated by a thorough review of multiple follow-ups, discharge summaries, the REDCap dataset, and consistent communication with the site's auditing personnel.

It is important to note that each CCRC site presents a unique makeup and internal procedures of data verification and extraction standards. Further, the amount of structured data collection, complexity of the medical record, and changes in personnel can all lead to increased discrepancy rates and decreased quality. For the Infant Tetralogy of Fallot cohort study, each CCRC institution employed its own unique organisational team makeup. As an example, some sites utilised cardiology fellows for data extraction, research coordinators for data entry ,and Principal Investigators for overall guidance and mentorship. Other sites consisted of smaller teams with the Principal Investigator extracting and entering data with support from one additional team member. Moreover, the cohort size was inconsistent with some sites contributing a disproportionate share of the total number of study records. The CCRC reports with timeliness and deadline adherence; however, it is important to note that inconsistencies amongst institutions can lead to bias, considerations for extensions, or unbalanced workloads.

Our data collection efforts included paper, hand-written, and electronic documentation review that spanned over a decade, presenting an additional layer of complexity. Older records were at times less thorough or detailed and required more effort to locate and secure. It is important to note that the majority of errors were clerical in nature, which is likely attributed to the manualised efforts, the span of medical records included, and the number of team members involved both at the internal site level and across all CCRC institutions. Similar datasets, national databases, and registries should consider a combination of manualised and automatic efforts when feasible to reduce the frequency of clerical errors.

While the two auditors received the same training and worked closely together, minor differences in interpretation are unavoidable. To mitigate variances, continuous check-ins were performed throughout the auditing process to increase consistency and transparency. For future auditing purposes, we suggest considering audit execution employing a single individual permitting bandwidth and feasibility. Lastly, as the auditors were not permitted to review files in person, it must be stated that this process relied

https://doi.org/10.1017/S1047951121000974 Published online by Cambridge University Press

heavily on the integrity of each institution's source data verification audit personnel. Thus, consistent training, communication, and Principal Investigator mentorship served as key components to the success of this data quality initiative.

"Regarding our reported per-patient major discrepancy rate, a few issues are worth noting. First, no errors in reporting the study's primary outcome of death were discovered. Thus, the primary outcome analysis was not affected by data errors in this study. Second, there were no differences in discrepancy rates between treatment groups, meaning any potential bias related to including discrepant data in analyses of secondary outcomes should be non-differential. Third, the vast majority of major discrepancies were related to inconsistencies in recorded dates outside of the strict data policy we established of +/− 3 days. We are reassured that if we had, instead, broadened this threshold to allow dates within 10 days, the per-patient major discrepancy rate would have decreased to 19%. Finally, to our knowledge, a per-patient major discrepancy rate is not a measure that has been reported by other cardiac collaboratives or registries to assess data quality."

Lastly, we would be remiss to not mention potential regulatory concerns. The process of de-identifying, uploading, and sharing files containing protected health information potentially allows opportunities for breaches of confidentiality. Such breaches are a potential concern with this, and any medical data auditing process. Careful attention needs to be paid to minimising protected health information exposures and inadvertent breaches of patient confidentiality.

Conclusion

The CCRC sought to establish a high-quality research dataset for use in our Infant Tetralogy of Fallot cohort study based on rates of accuracy, completeness, and timeliness. While the collaborative will consider opportunities to conduct an in-person audit in its future, we believe our results demonstrate that remote, non-automatised audits of multicentre retrospective cohort studies can produce high-quality, reliable data. Results from the collaborative's first source data verification audit demonstrate high-quality accuracy with no evidence of omission or adjustments of the key auditing data variables. Study findings further demonstrate source data verification audits can identify integrity and quality concerns that can be applied to various data quality assurance studies. These processes may be particularly useful during the current COVID-19 pandemic.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/S1047951121000974.

Acknowledgements. The authors recognise the work of all data collection and auditing team members at the participating CCRC sites whose efforts contributed to the data quality.

Financial support. Financial support for this research was derived, in part, from the Kennedy Hammill Pediatric Cardiac Research Fund, the Liam Sexton Foundation, and A Heart Like Ava.

Conflicts of interest. None.

References

1. Petit CJ, Qureshi AM, Glatz AC, et al. Comprehensive comparative outcomes in children with congenital heart disease: the rationale for the Congenital Catheterization Research Collaborative. Congenit Heart Dis 2019; 14: 341–349. doi: 10.1111/chd.12737.
2. Petit CJ, Glatz AC, Qureshi AM, et al. Outcomes after decompression of the right ventricle in infants with pulmonary atresia with intact ventricular septum are associated with degree of tricuspid regurgitation: results from the Congenital Catheterization Research Collaborative. Circ Cardiovasc Interv 2017; 10: e004428. doi: 10.1161/CIRCINTERVENTIONS.116.004428. Erratum in: Circ Cardiovasc Interv. 2017 Jun;10(6).
3. Petit CJ, Gao K, Goldstein BH, et al. Relation of aortic valve morphologic characteristics to aortic valve insufficiency and residual stenosis in children with congenital aortic stenosis undergoing balloon valvuloplasty. Am J Cardiol 2016; 117: 972–979. doi: 10.1016/j.amjcard.2015.12.034
4. Glatz AC, Petit CJ, Goldstein BH, et al. A Comparison Between Patent Ductus Arteriosus Stent and Modified Blalock-Taussig Shunt as Palliation for Infants with Ductal-Dependent Pulmonary Blood Flow: insights from the Congenital Catheterization Research Collaborative. Circulation 2018; 137: 589–601. 10.1161/CIRCULATIONAHA.
5. Goldstein BH, Petit CJ, Qureshi AM, et al. Comparison of Management Strategies for the Neonate with Symptomatic Tetralogy of Fallot: A Study from the Congenital Cardiac Research Collaborative. J Am Coll Cardiol 2021; 77: 1093–1106.
6. Giganti MJ, Shepherd BE, Caro-Vega Y, et al. The impact of data quality and source data verification on epidemiologic inference: a practical application using HIV observational data. BMC Public Health 2019; 19: 1748. doi: 10.1186/s12889-019-8105-2.
7. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG Research Electronic Data Capture (REDCap): a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009; 42: 377–381. doi: 10.1016/j.jbi.2008.08.010.
8. Daniels K, Lawson E Enhancing Clinical Librarian Work Processes and Communicating Service Impact with REDCap and Data Visualization. Med Ref Serv Q 2019; 38: 187–196. doi: 10.1080/02763869.2019.
9. Gaies M, Donohue JE, Willis GM, et al. Data integrity of the Pediatric Cardiac Critical Care Consortium (PC4) clinical registry. Cardiol Young 2016; 26: 1090–1096. doi: 10.1017/S1047951115001833.
10. Houston L, Probst Y, Humphries A. Measuring data quality through a source data verification audit in a clinical research setting. Stud Health Technol Inform 2015; 214: 107–113.
11. Messenger JC, Ho KK, Young CH, et al. NCDR Science and Quality Oversight Committee Data Quality Workgroup. The National Cardiovascular Data Registry (NCDR) Data Quality Brief: the NCDR Data Quality Program in 2012. J Am Coll Cardiol 2012; 60: 1484–1488. doi: 10.1016/j.jacc.2012.07.020.
12. Gaies M, Cooper DS, Tabbutt S, et al. Collaborative quality improvement in the cardiac intensive care unit: development of the Paediatric Cardiac Critical Care Consortium (PC4). Cardiol Young 2015; 25: 951–957. doi: 10.1017/S1047951114001450.
13. Overman DM, Jacobs ML, O'Brien JE, et al. Ten Years of Data Verification: the Society of Thoracic Surgeons Congenital Heart Surgery Database Audits. World J Pediatr Congenit Heart Surg 2019; 10: 454–463. doi: 10.1177/2150135119845256.
14. Andersen JR, Byrjalsen I, Bihlet A, et al. Impact of source data verification on data quality in clinical trials: an empirical post hoc analysis of three phase 3 randomized clinical trials. Br J Clin Pharmacol 2015; 79: 660–668. doi: 10.1111/bcp.12531.
15. Mealer M, Kittelson J, Thompson BT, et al. Remote source document verification in two national clinical trials networks: a pilot study. PLoS One 2013; 8: e81890. doi: 10.1371/journal.pone.0081890.

https://doi.org/10.1017/S1047951121000974 Published online by Cambridge University Press